

Instructions for Using the ICH_ICSR.DTD to Prepare E2B Compliant Individual Case Safety Reports for Electronic Transmission

Ver 1.3 1 December, 1997

The ICH E2B Expert Working Group (EWG) was formed to standardize the data elements for transmission of Individual Case Safety Reports (ICSR) for both pre, and post-approval periods and cover both adverse drug reaction and adverse event reports. The ICH M2 EWG helped facilitate the completion of the E2B data definitions by creating a relational view, attribute list, and Structured Generalized Markup Language (SGML) Document Type Definition (DTD) that allows for structured data sets of all types of individual case safety reports to be transported from database-to-database.

This document provides directions for preparing a structured data set of E2B compliant individual case safety reports with the use of the ICH_ICSR.DTD. The first part of the document provides instructions for creating a SGM data file that can be used for electronic transmission. The second part provides instructions to use a SGML parser, such as Perl, to test the data set for syntax errors and to extract the structured information to populate adverse events databases, like the FDA's Adverse Events Reporting System (AERS) ORACLE database. **The instructions to use Perl have been provided for guidance and as an example. ICH M2 does not mandate the use of Perl to parse the SGML data.**

1. Instructions for developing a SGM data file that is compliant with the ICH_ICSR.DTD

This section explains some basic steps to help prepare a SGM data set that is compliant with the ICH_ICSR.DTD and can be used for the electronic transmission of structured individual case safety data.

The E2B Step 4 document, the M2 Relational view diagram (m2relv22.pdf), and the spreadsheet listing the field attributes agreed between ICH E2B and M2 (E2BS4V34.PDF) along with DTD Descriptors can be used as a reference while entering data to verify the accuracy of the DTD descriptor and E2B field definitions. Typically, SGML editing tools can be used to prepare a data set that is specific to an SGML DTD. The SGML tools check for valid SGML rules and provide efficient features, like automatic insertion of start and end tags for the easy preparation of data. However, in the absence of SGML editing tools, a word processor or simple text editor can also be used in the following manner:

- Open ICH_ICSR.DTD with a word processor or text editor
- Open a new text document and start creating an SGM file by defining tags for each DTD element

For example, in the new document, define tag as <ICHICSR> for element defined in DTD as <!ELEMENT ichicsr - - (safetyreport*) >.
- Adverse event data must be entered after the open tag of the appropriate descriptors defined in the new text file. To verify that data is populated appropriately, check the E2B Step 4 document and M2 field attribute spreadsheet (E2BS4V34.XLS) for field definitions, title, field length, field value, and descriptor name

For example, in the populated SGML file – smallts2.sgm, the value “US” is entered after the field descriptor <PRIMARYSOURCECOUNTRY>, as <PRIMARYSOURCECOUNTRY>US. This value was entered after verifying that the tag <PRIMARYSOURCECOUNTRY> referred to the E2B field A 1.1, identification of the country where the reaction/event, with DTD descriptor primarysourcecountry, field length of 2AN, and a country code compliant with ISO3166, US.
- Make sure that there is an end tag for every start tag

For example, <tagname> must be followed by </tagname>...

<PRIMARYSOURCECOUNTRY>US</PRIMARYSOURCECOUNTRY>

Instructions for Using the ICH_ICSR.DTD to Prepare E2B Compliant Individual Case Safety Reports for Electronic Transmission

Ver 1.3 1 December, 1997

- To make a data file readable to a human, carriage returns can be inserted after each end tag. However, while testing the DTD with SGM test files, a certain type of end tag error was generated. The reason for the error was attributed to the carriage return characters that were present after some of the end tags. Since the nature of the error was ambiguous in nature, the error was corrected by deleting all carriage returns or blank characters following the end tags.

For example,

```
<PRIMARYSOURCECOUNTRY>US</PRIMARYSOURCECOUNTRY><OCCURCOUNTRY>US</OCCURCOUNTRY><TRANSMISSIONDATEFORMAT>102</TRANSMISSIONDATEFORMAT><TRANSMISSIONDATE>19970101</TRANSMISSIONDATE>
```

Note: To produce an error free data file, carriage returns can also be removed from only those lines that are identified by the parser to have an end tag error.

- Although you might need to delete all carriage returns between an end tag and the following start tag, you can have carriage returns in the middle of large text fields

For example, between start and end tag of a 200 AN text field you can enter text with carriage return.

```
<NULLIFICATIONREASON>invalid information regarding ...  
...the drug reaction</NULLIFICATIONREASON>
```

- The repeatable fields and sections (designated by a double line box in the M2 field attribute spreadsheet - E2BS4V34.XLS) are populated by copying and pasting the DTD descriptors, the required number of times and then populating them with data, repeatedly.

Note: Each section that is repeated, must begin with the section name (<REPORTDUPLICATE >) and end with the corresponding end tag (back slash before the same title </REPORTDUPLICATE >).

For example,

```
<REPORTDUPLICATE><REPORTDUPLICATESQ>001<DUPLICATESOURCE>bones  
pharmaceuticals</DUPLICATESOURCE><DUPLICATENUMB>op89</DUPLICATENUMB>  
</REPORTDUPLICATESQ><REPORTDUPLICATESQ>002<DUPLICATESOURCE>white  
labs</DUPLICATESOURCE><DUPLICATENUMB>87po</DUPLICATENUMB></REPORTDUPLICATESQ></REPORTDUPLICATE>
```

- For further illustration, the following scripts have been provided as an example of a DTD definition and the corresponding syntax of a populated SGM file.

- **Part of the ICH_ICSR.DTD**

```
<!ELEMENT ichicsr - - ( safetyreport* ) >  
<!ELEMENT safetyreport - - ( safetyreportid ) >  
<!ELEMENT safetyreportid - - ( #PCDATA,safetyreportsq+ ) >  
<!ATTLIST safetyreportid %ope.rec; >  
<!ELEMENT safetyreportsq - - ( #PCDATA,  
    (primarysourcecountry? &  
    occurcountry? &  
    ...
```

Instructions for Using the ICH_ICSR.DTD to Prepare E2B Compliant Individual Case Safety Reports for Electronic Transmission

Ver 1.3 1 December, 1997

<!ATTLIST safetyreportsq %ope.rec; >

- **Sample SGM data file**

```
<ICHICSR><SAFETYREPORT><SAFETYREPORTID><SAFETYREPORTSQ><PRIMARYSOURCECOUNTRY>US</PRIMARYSOURCECOUNTRY><OCCURCOUNTRY>US</OCCURCOUNTRY>...
```

...

```
<REPORTDUPLICATE><REPORTDUPLICATESQ>001<DUPLICATESOURCE>bones pharmaceuticals</DUPLICATESOURCE><DUPLICATENUMB>op89
```

```
</DUPLICATENUMB>></REPORTDUPLICATESQ><REPORTDUPLICATE>
```

```
<REPORTDUPLICATESQ>002<DUPLICATESOURCE>white labs</DUPLICATESOURCE><DUPLICATENUMB>87po</DUPLICATENUMB>
```

```
</REPORTDUPLICATESQ></REPORTDUPLICATE>
```

...

```
</SAFETYREPORTSQ></SAFETYREPORTID></SAFETYREPORT></ICHICSR>
```

- Once all the E2B compliant information is entered in the appropriate sections, the document must be saved as, filename.SGM (SMALLTS2.SGM).

The E2B compliant data must be tested for extraction and syntax errors, prior to transmission. Data can be extracted from the populated SGML file, SMALLTS2.SGM, by executing an SGML parser program, like the Perl SGML program (pe2b.pl). The parsing of the SGML file will result in the generation of a SQL script file (smallts2.sql). This SQL script can then be used to populate an ORACLE relational schema that is E2B compliant.

- If the parser generates end tag errors, it is probably due to carriage returns being inserted in inappropriate places. As stated before, please make sure that there are no extra carriage returns or other characters following those end tags.
- Make sure the SGM syntax refers to the ICH_ICSR.DTD in the SGM file (the first and last tags must be <ICHICSR> and </ICHICSR>, respectively).

Instructions for Using the ICH_ICSR.DTD to Prepare E2B Compliant Individual Case Safety Reports for Electronic Transmission

Ver 1.3 1 December, 1997

2. Instructions for Installing Perl and Extracting SGML Data

The instructions to use Perl have been provided for guidance and as an example. ICH M2 does not mandate the use of Perl to parse the SGML data.

As mentioned in section 1, after creating the SGM file, it must be tested for syntax and data extraction errors. This can be achieved by parsing data out of the SGM file with the use of a parser program. This part of the document provides instructions for installing Perl and running the extraction programs. As mentioned before, the instructions to use Perl has been provided for guidance and as an example. ICH-M2 does not mandate the use of Perl to parse the SGML data.

The FDA's development team that modified, tested, and validated the ICH_ICSR.DTD used the Perl SGML parser to extract data from the data file. Perl is an interpreter language optimized for scanning arbitrary text files, like SGML data files. Perl allows for parsing of SGML data, extracting information into SQL scripts, and printing reports based on extracted information.

The following instructions can also be used as a reference for using other SGML parsers:

Installation of Perl

- The Perl SGML parser application is available on the Internet. The Perl application files can be obtained by downloading three zip files, perl.zip, sp.zip, and perlsgml.zip from the following Internet http addresses:

perl5a1.zip: <http://www.perl.com>
or <http://www.winternet.com/~plp/perl5/>

sp1_1_1.zip: <http://www.jclark.com/sp/>

Perlsgml.zip: <http://www.oac.uci.edu/indiv/ehood/>

Once the zip files have been downloaded the following steps must be completed:

- In MS-DOS mode, unzip the perl.zip, sp.zip, and perlsgml.zip files using pkunzip. Unzipping these files will result in the creation of several sub-directories containing Perl application files
- Move the following files to the Perl-5.000 sub-directory:
 - **sgml.pl:** This is a Perl program, required by the parser program pe2b.pl, custom developed to extract data from the populated E2B/M2 SGML files. (Copy sgml.pl from /lib, a sub-directory created by unzipping perlsgml.zip.)
 - **newgetop.pl:** This is a Perl program required for the execution of the pe2b.pl program. (Copy this file from /lib sub-directory, created by unzipping the Perl zip file.)
 - **nsgmlnorm.exe:** This is an executable program required to run the pe2b.pl program. (Copy this from /sp sub-directory after unzipping the spzip file.)
- Now, copy the following files to the Perl-5.000 sub-directory:

Instructions for Using the ICH_ICSR.DTD to Prepare E2B Compliant Individual Case Safety Reports for Electronic Transmission

Ver 1.3 1 December, 1997

- **pe2b.pl:** This is the Perl program developed by the FDA's development team to parse the SGML document and produce the SQL script file.
- **ich_icsr.dtd:** This is the ICH E2B/M2 SGML DTD that is compliant with the E2B Step 4 document and the M2 Attribute list, version 3.4 (E2BS4V34.PDF).
- **ich_icsr.dcl:** This is a SGML declaration file that must be used along with the DTD.
- **Smallts2.sgm:** This is a SGML test file that is populated with sample fictitious E2B compliant data extracted from a Medwatch form.

Extraction of SGML data into relational database tables

Once the Perl SGML parser application has been installed and the appropriate files have been placed in the Perl-5000 sub-directory, data from the populated SGML file, SMALLTS2.SGM, can be extracted by the execution of the command **Perl pe2b.pl -f SMALLTS2>SMALLTS2.SQL**

The execution of the parser program generates a SQL script, smallts2.sql that can be used to load data into a relational database, like ORACLE. At the FDA, the SQL script SMALLTS2.SQL was run through an ORACLE SQL load procedure to populate the E2B schema in the AERS database. Once the E2B compliant adverse events information is loaded into a relational database, the data can then be further processed into the normalized and typically more complex relational data model of the receiving database.

Similarly, to extract data from relational database, the normalized data model must be mapped to the E2B model, and extraction routines must be written to process data that can then be used to prepare a SGM data file that is compliant with the ICH_ICSR.DTD.